BIOLOGY RELATED INFORMATION STORAGE KIT

User Manual

Investigator Services

The Daley Lab

12/1/2010

TABLE OF CONTENTS

Getting Started 3
User Logins 3
Generating Analysis Files 3
Generate Raw Genotype Analysis Files 3
Generate Analysis Files/Request Data File 5
My Data Requests 8
Data Request Approvals 8
Search 8
General (All) Database Search 9
SNP Search 9
Gene Search 10
KEGG Pathway Search 12
Mapping Features 12
Gene to SNP: 13
SNP to Gene: 14
Gene to Pathway: 16
Pathway to Gene: 17
Association 19
Make Pathway 20
Gene-Gene Interaction 20
My Results 21
Site Map 21
Admin Settings 22
Show/Hide SNPs/Genes 22
Update Gene Links 22
Update Gene Alias 22
Update KEGG Pathways 22

INVESTIGATOR SERVICES

GETTING STARTED

After deploying Biology Related Information Storage Kit (BRISK) or Investigator Services onto the server, you will need to populate the database backend with data before all the features can be properly used. The database can be populated manually, or through the use of the import tool in Sample-base Laboratory Information Management System (SLIMS).

USER LOGINS

In order to use Investigator Services, you must have an account with permission to use Investigator Services. New user logins can be created manually through the database, or through user accounts management in SLIMS.

GENERATING ANALYSIS FILES

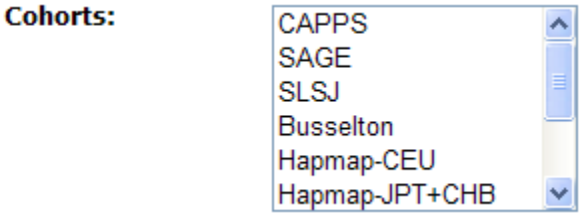
There are actually 5 different options under the "Generate Analysis Files" header. What appears under that menu depends greatly on what sort of access credentials is linked to your account. The 5 different options are:

- Generate Raw Genotype Analysis Files [Full data access]
- Generate Analysis Files [Full data access]
- My Data Requests [Partial data access]
- Request Data File [Partial data access]
- Data Request Approvals [Data owners]

GENERATE RAW GENOTYPE ANALYSIS FILES

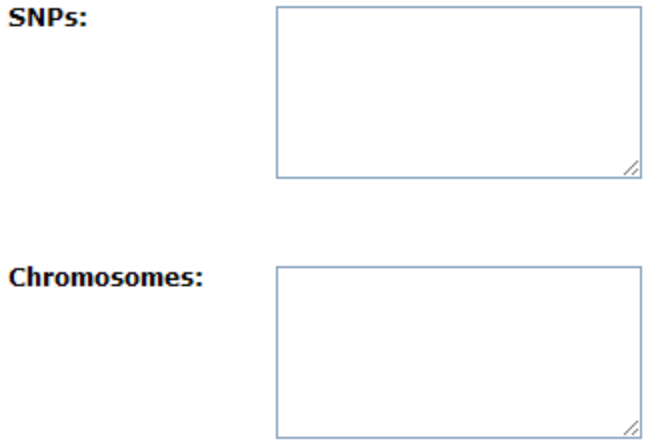
The feature generates raw genotype analysis files, in Linkage file format, for the user to use.

You can select which dataset you want to generate the analysis files from. By holding down CTRL and then clicking on each desired genotyping run, you can select multiple datasets to generate the analysis files from.



You can specify the cohorts you would like to see in the analysis file. By holding down CTRL and then clicking on each desired cohort, you can select multiple cohorts to include in the analysis file.

You can specify a specific phenotype you want to include in the analysis file.



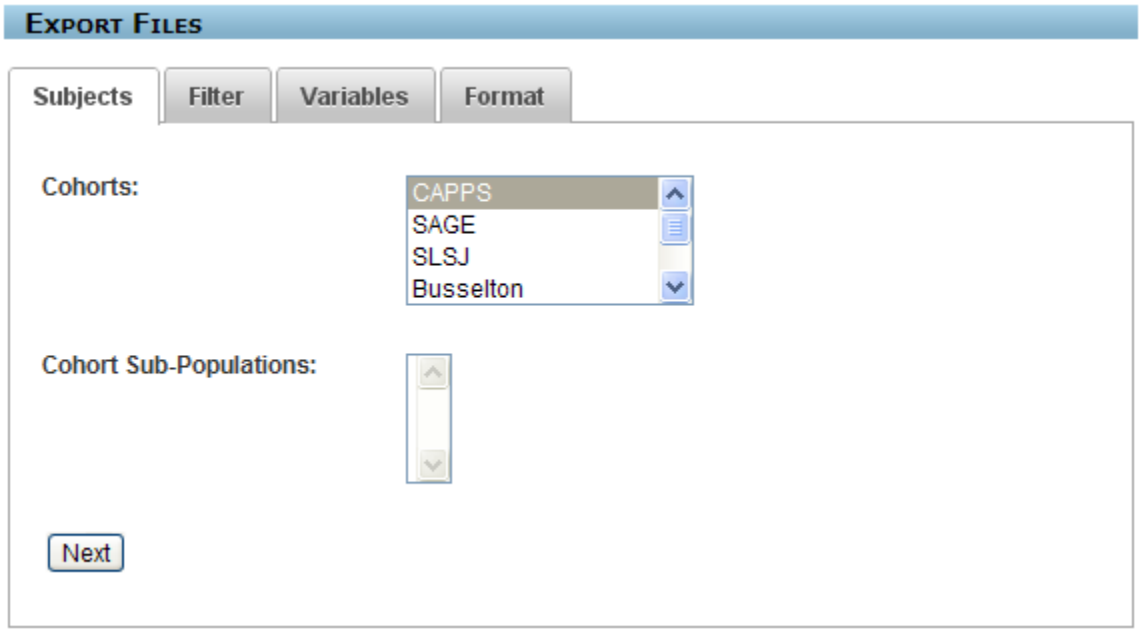
In the boxes, pictured above, you can freely type markers into the two textboxes (one for SNPs, and one for chromosomes) to include in the analysis files. There should only be one SNP or chromosome per line.

Here, pictured above, you can fill out the email address you would like to be sent the analysis files.

GENERATE ANALYSIS FILES/REQUEST DATA FILE

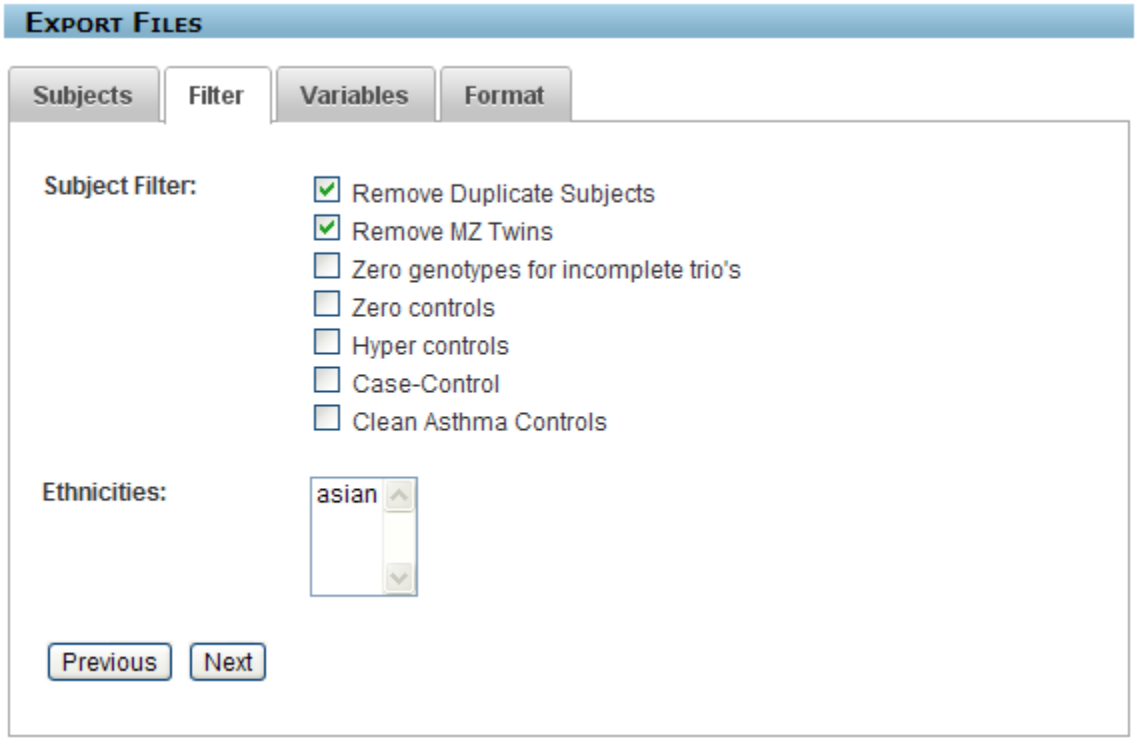
This feature allows you to generate/request analysis files in a file format of your choice. The options are split into different categories to allow users to easily change the options in the category they are most interested in.

Subjects



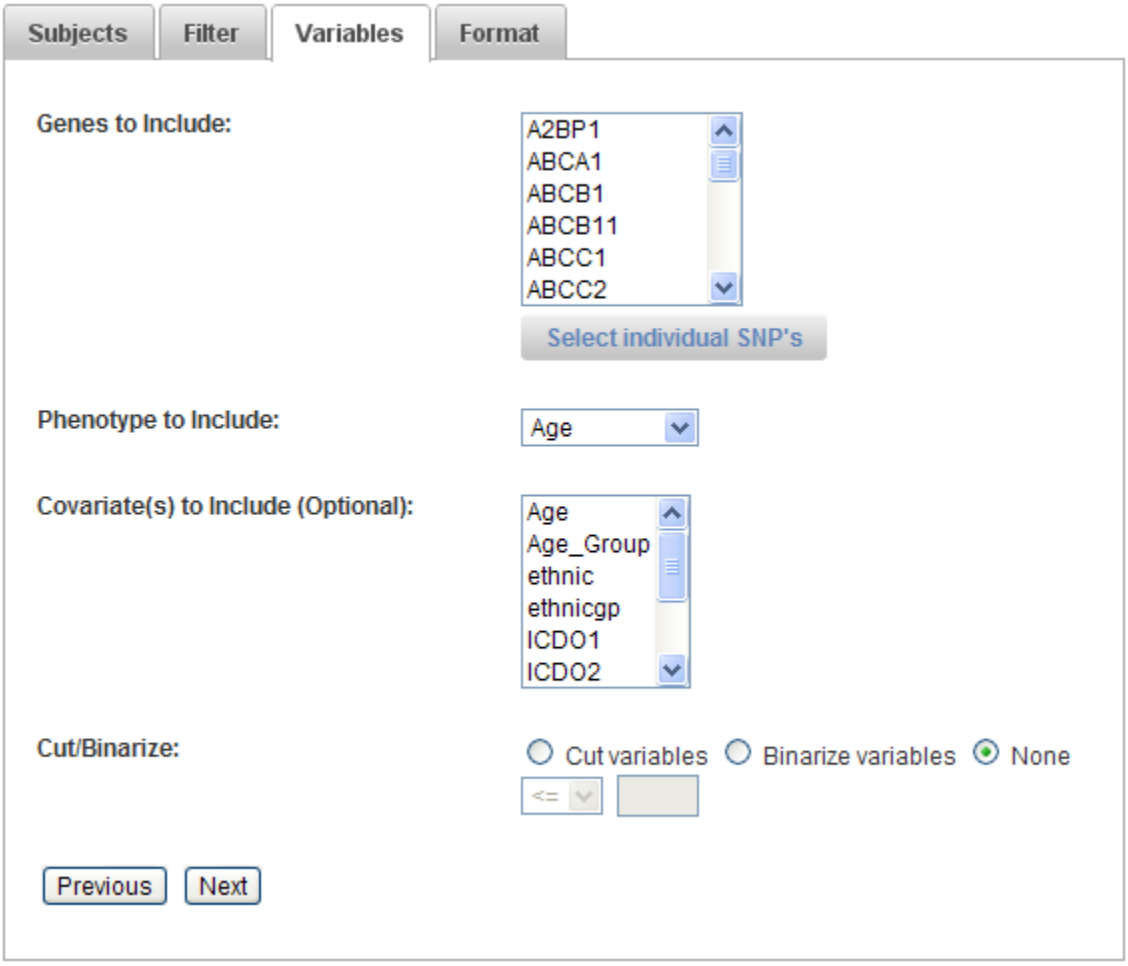
The options under this tab allow you to choose what cohort and subpopulations you would like to include in the analysis file. Each box is multi-selectable; by holding down the CTRL key, you can click and select multiple cohorts and sub-cohorts to include in the analysis.

Filter



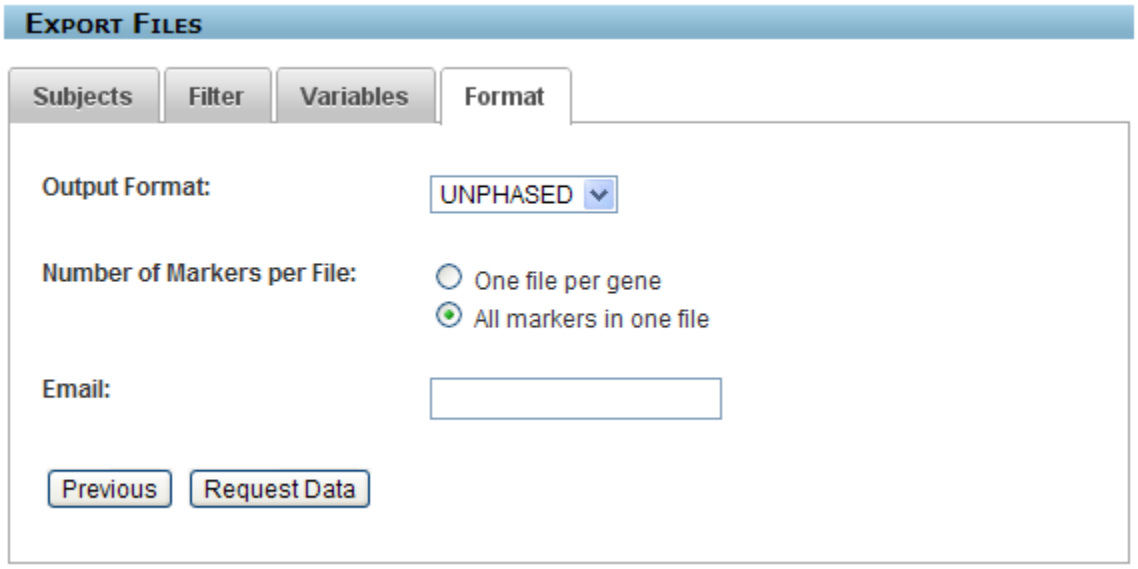
The options under this tab allow you to finely filter out unwanted data in the analysis file. You are able to specify what to exclude from the analysis file through simple checkboxes. Furthermore, you are able to further fine-tune the cohort by including only certain ethnicities in the analysis file. The ethnicities box is multi-selectable, to select several ethnicities to include, hold CTRL and click on each ethnicity that needs to be included.

Variables



The options under this table allow a user great control of what to include in the analysis file. You are able to specify which genes you would like to see in the analysis file – you are even able to specify what SNPs from each gene they would like to include. You can choose what phenotype you want your analysis file to look at. Additionally, you are able to select multiple covariates to include. There is also an input box is available if you want to cut/binarize a variable.

Now finally, you can specify the output format of the file. Currently UNPHASED, PLINK, fbat, and SAGE file formats are supported. Next you have to choose whether you want each gene in its own file, or one large file containing all the genes.



This is where the difference between "requesting" and "generating" a file happens. If you are generating an analysis file, you will be able to specify an email address for the server to email your files. However, if you were requesting for an analysis file, you will have to wait for approval by the proper people before you will have access to your analysis file.

My Data Requests

This section allows you to view all the pending data requests you have available. You are able to see the details of each request, along with approval progress. If you want to cancel the data request, this can be done on this page as well.

Data Request Approvals

If you see this option in the menu, that means you are a data owner. On this page, you can view all the data requests that contain data owned by you. You can approve or deny data requests through a drop down menu. Additional comments can be added to the data request through the little textbox.

Search

This set of features, allows you to search through the database for genes, SNPs, and pathways.

General (All) Database Search

By clicking on the "Search" heading, in the main menu bar, you are brought to a search box that allows you search through gene, SNP, and pathways. The program will try to find all mentions of what is inputted into the textbox. If you are looking for multiple items, please type one item per line.

SNP Search

This search will search through the database for SNPs you inputted. The search box takes one RS number or SNP alias per line. Additionally, you may want to click on "Show All" to view all the SNPs in the database.

Advanced Search

This gives you further expansion for your search criteria. You can specify a certain gene, chromosome, or pathway you want to look at and the results will show you SNPs that match any of the search criteria.

Gene Criteria

The search will return all the SNPs in the specified genes, and all the SNPs generated by another search box; it takes a union of the results, and the SNPs in the genes. Multiple genes can be specified by using multiple lines and typing one gene on each line.

Chromosome Criteria

The search will return all the SNPs in the specified chromosomes, and all the SNPs generated by another search box; it takes a union of the results, and the SNPs in the chromosomes. Multiple chromosomes can be specified by using multiple lines and typing one chromosome on each line.

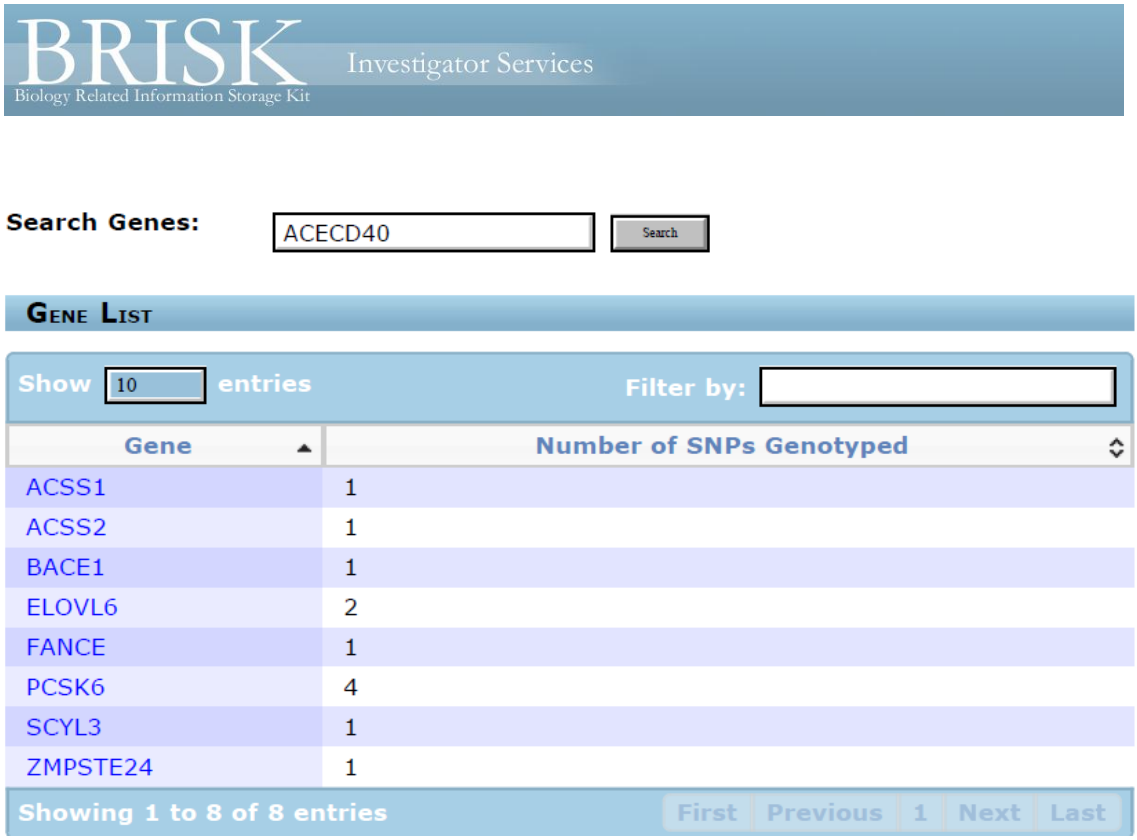
KEGG Pathway Criteria

The search will return all the SNPs in the specified pathways, and all the SNPs generated by another search box; it takes a union of the results, and the SNPs in the pathways. Multiple pathways can be specified by using multiple lines and typing one pathway on each line.

Gene Search

This search will search through the database for genes you inputted. The search box takes one gene name per line. Additionally, you may want to click on “Show All” to view all the genes in the database

This search returned the following



Note that none of the gene names match the search criteria but there are partial matches in the gene alias.

Advanced Search

This gives you further expansion for your search criteria. You can specify a certain chromosome, KEGG Pathway, or publications you want to look at and the results will show you genes that match any of the search criteria.

Chromosome Criteria

The search will return all the genes in the specified chromosomes, and all the genes generated by another search box; it takes a union of the results, and the genes in the chromosomes. Multiple chromosomes can be specified by using multiple lines and typing one chromosome on each line.

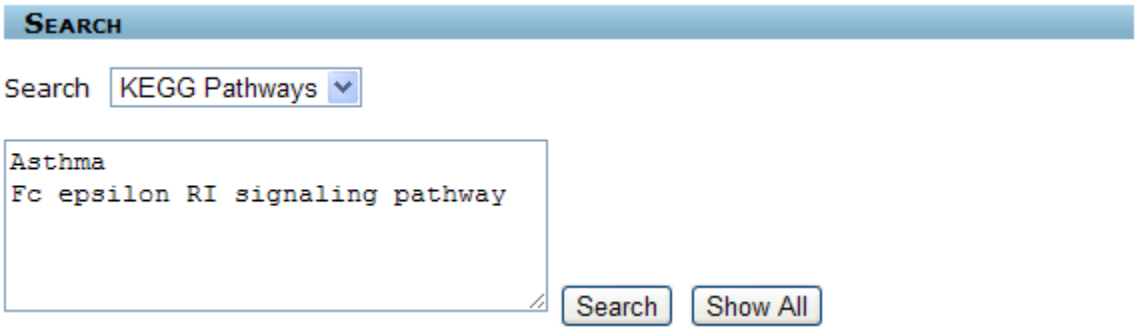
KEGG Pathway Criteria

The search will return all the genes in the specified pathways, and all the genes generated by another search box; it takes a union of the results, and the genes in the pathways. Multiple pathways can be specified by using multiple lines and typing one pathway on each line.

Publication/Pending Publication Criteria

The search will return all the genes that were in a specific publication by the lab, and all the genes generated by another search box; its takes the union of the results, and the genes in the publication.

KEGG PATHWAY SEARCH



This search will search through the database for SNPs you inputted. The search box takes one RS number or SNP alias per line. Additionally, you may want to click on "Show All" to view all the SNPs in the database.

Advanced Search

This gives you further expansion for your search criteria. You can specify a certain gene, chromosome, or pathway you want to look at and the results will show you SNPs that match any of the search criteria.

Gene Criteria

The search will return all the pathways containing the specified genes, and all the pathways generated by another search box; it takes a union of the results, and the pathways containing the genes. Multiple genes can be specified by using multiple lines and typing one gene on each line.

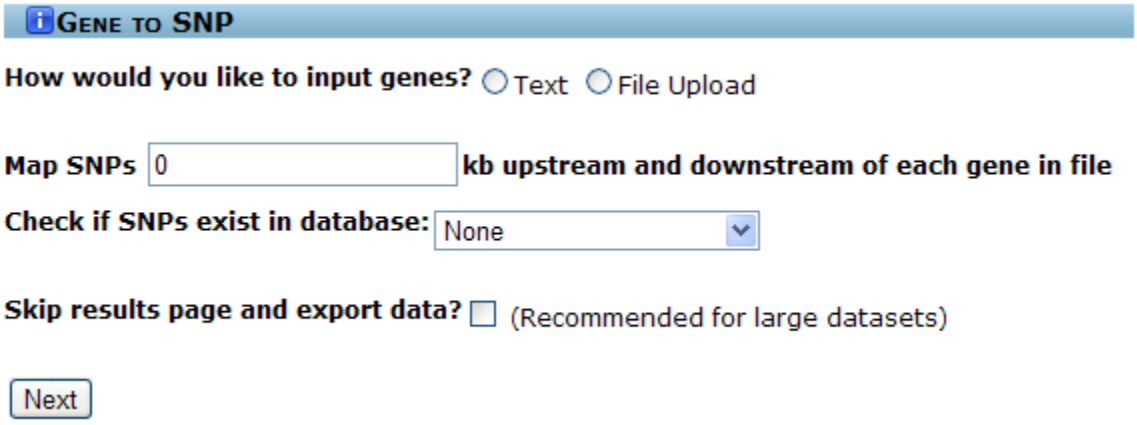
SNP Criteria

The search will return all the pathways containing the specified SNPs, and all the pathways generated by another search box; it takes a union of the results, and the pathways containing the SNPs. Multiple SNPs can be specified by using multiple lines and typing one SNP on each line.

MAPPING FEATURES

These features allow a user to easily transition from one set of data to another set of related data.

GENE TO SNP:

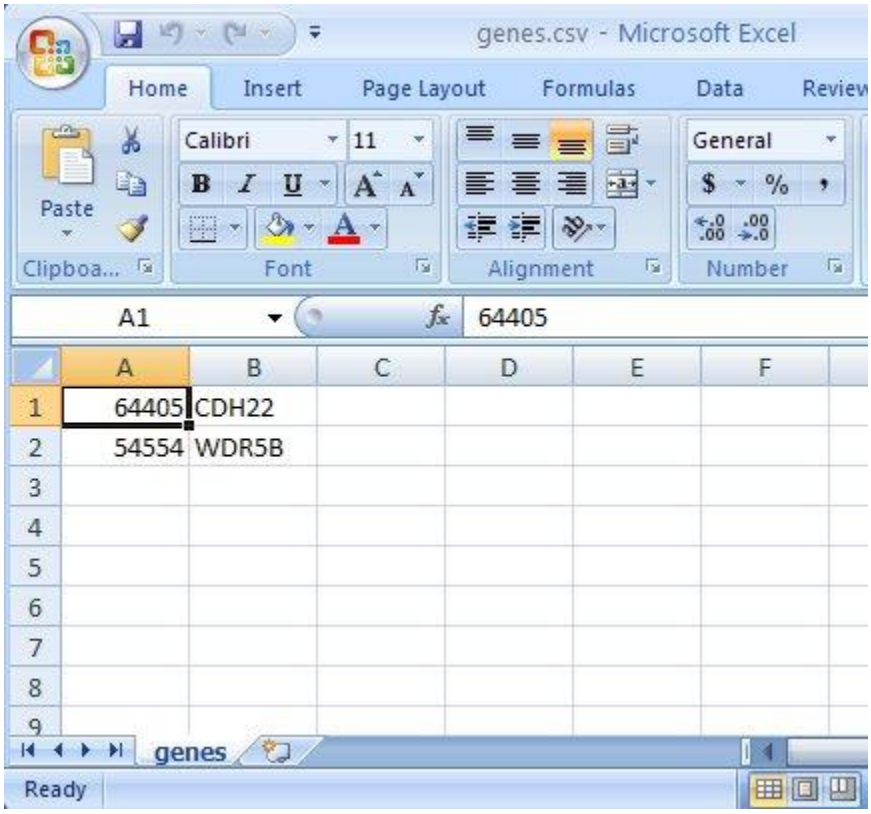

GENE TO SNP
How would you like to input genes? Text File Upload
Map SNPs 0 kb upstream and downstream of each gene in file
Check if SNPs exist in database: None
Skip results page and export data? (Recommended for large datasets)
Next

This feature queries the NCBI database for all the SNPs that fall within the chromosome range spanned by the specified genes. The results are displayed in a table format and can be easily exported to a file. Note that this feature takes the same gene list as the one used by Gene to Pathway.

How to use:

1) This feature takes a gene list through a file or a textbox.
    a. The text input requires that the user type their gene list into a text box provided on the page. Each line must contain a "NCBI gene ID" followed by a space and then the corresponding gene name for that ID.
        i. For example:
            64405 CDH22
            54554 WDR5B
    b. File upload requires that the user upload a preformatted file containing the gene list. The gene list should be a XLS, CSV, or tab-delimited TXT file. The file should contain two columns:
        i. First column with the NCBI Gene ID
        ii. Second column with the Gene Name

FIGURE 1: SCREENSHOT SHOWING A CSV GENE LIST

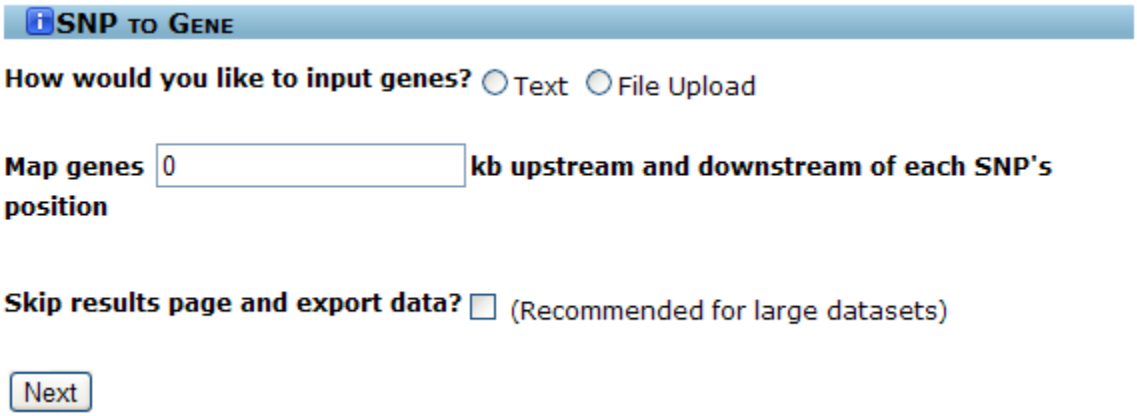


2) After providing the gene list, either through uploading a file or typing the gene list, you can click "Next" to continue. If you would like to change the other options, please refer to the next section for a description of each option.
3) The results page contains a table displaying the data returned. The table has sorting, filtering, and paging options. Below the table is a link that allows you to export the data into a file with the filename and delimiter of your choice.

Additional Options:

- The user is able to specify an extension (in kilobase) upstream and downstream of the chromosome range spanned by the specified gene to map SNPs.
- A list of SNPs can also be uploaded to do a comparison; the results table will indicate which SNPs from the query were also found in the uploaded SNPs list.
- The results from the NCBI query can be exported to a file immediately, bypassing the results table.

SNP TO GENE:


SNP TO GENE
How would you like to input genes? Text File Upload
Map genes 0 kb upstream and downstream of each SNP's position
Skip results page and export data? (Recommended for large datasets)
Next

This feature queries the NCBI database for the genes that the SNP falls on. The results are displayed on an interactive table with sorting, paging, and filtering capabilities.

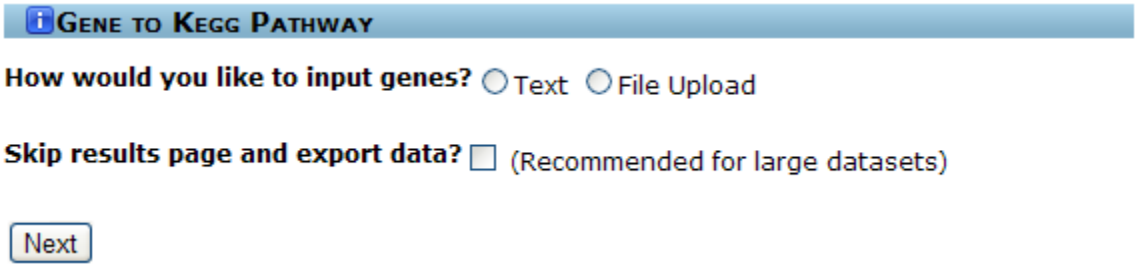
How to use:

1) This feature can take a SNP list through a file or a textbox.
    a. The text input requires that the user type their SNP list into the text box provided on the site. Each line must contain a RS number with an "rs" prefix.
        i. For example:
            rs1067
            rs4165
    b. File upload requires that the user upload a preformatted file containing the SNP list. The SNP list should be a TXT or XLS file. Each line in the file should contain a RS number – like the above example.
2) After providing the SNP list, you can click "Next" to continue. If you would like to change the other options, please refer to the next section for more information.
3) The results page contains a table displaying the data returned. The table has sorting, filtering, and paging options. Below the table is a link that allows you to export the data into a file with the filename and delimiter of your choice.

Additional Options:

- The user is able to specify an extension (in kilobase) upstream and downstream of each SNP's position to map genes.
- The results from the NCBI query can be exported to a file immediately, bypassing the results table.

GENE TO PATHWAY:


GENE TO KEGG PATHWAY
How would you like to input genes? Text File Upload
Skip results page and export data? (Recommended for large datasets)
Next

This feature takes a list of genes and returns all the pathways that the specified genes are in. The results are displayed on an interactive table with sorting, paging, and filtering capabilities. Note that this feature takes the same gene list as the one used by "Gene to SNP"

How to use:

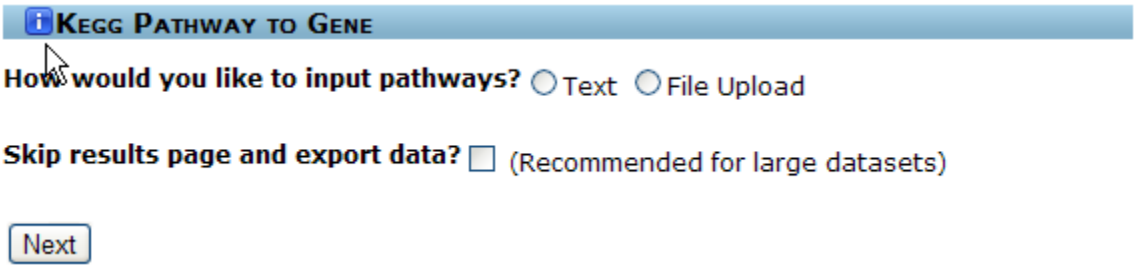
1) This feature can take a gene list through a file or a textbox.
    a. The text input requires that the user type their gene list into the text box provided on the page. Each line must contain a "NCBI gene ID" followed by a space and then the corresponding gene name for that ID.
        i. For example:
            64405 CDH22
            54554 WDR5B
    b. File upload requires that the user upload a preformatted file containing the gene list. The gene list should be a XLS, CSV, or tab-delimited TXT file. The file should contain two columns:
        i. First column with the NCBI Gene ID
        ii. Second column with the Gene Name
2) After providing the gene list, either through uploading a file or typing the gene list, you can click "Next" to continue. If you would like to change the other options, please refer to the next section for a description of each option.

3) The results page contains a table displaying the data returned. The table has sorting, filtering, and paging options. Below the table is a link that allows you to export the data into a file with the filename and delimiter of your choice.

Additional Options:

- The results from the query can be exported to a file immediately, bypassing the results table.

PATHWAY TO GENE:


KEGG PATHWAY TO GENE
How would you like to input pathways? Text File Upload
Skip results page and export data? (Recommended for large datasets)
Next

This feature takes a list of Kegg pathway names and returns all the genes that are in that pathway. The results are displayed on an interactive table with sorting, paging, and filtering capabilities.

How to use:

1) This feature can take a Kegg pathway list through a file or a textbox.
    a. The text input requires that the user type their Kegg pathway names into the textbox provided on the page. Each line must contain a Kegg pathway name.
        i. For example:
            Endocytosis
            Asthma
    b. File upload requires that the user upload a preformatted file containing the pathway list. The pathway list should be a TXT or XLS file. Each line in the file should contain a pathway name – like the above example.
2) After providing the pathway list, you can click “Next” to continue. If you would like to change the other options, please refer to the next section for more information.
3) The results page contains a table displaying the data returned. The table has sorting, filtering, and paging options. Below the table is a link that allows you to export the data into a file with the filename and delimiter of your choice.

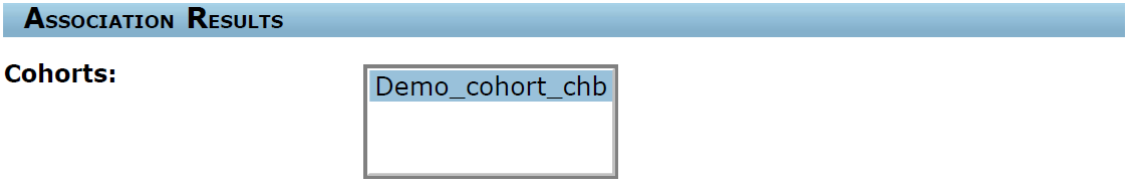
Additional Options:

- The results from the query can be exported to a file immediately, bypassing the results table.

ASSOCIATION

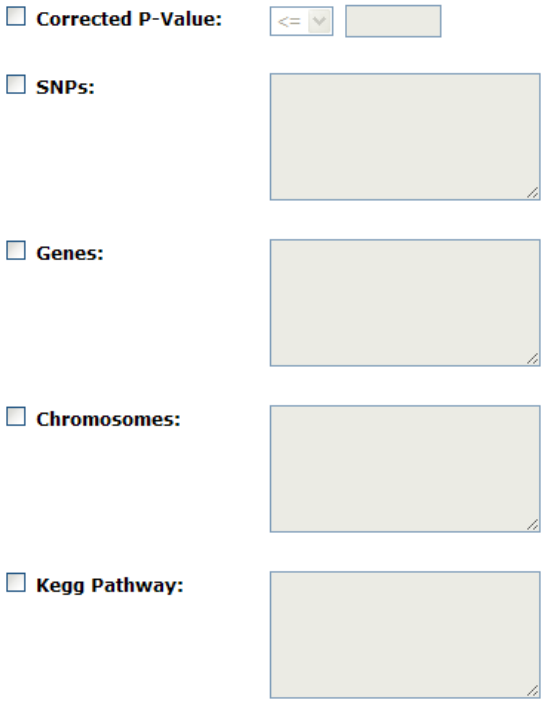
This feature allows you to see the association of the SNPs within a gene with a specific phenotype.

Usage

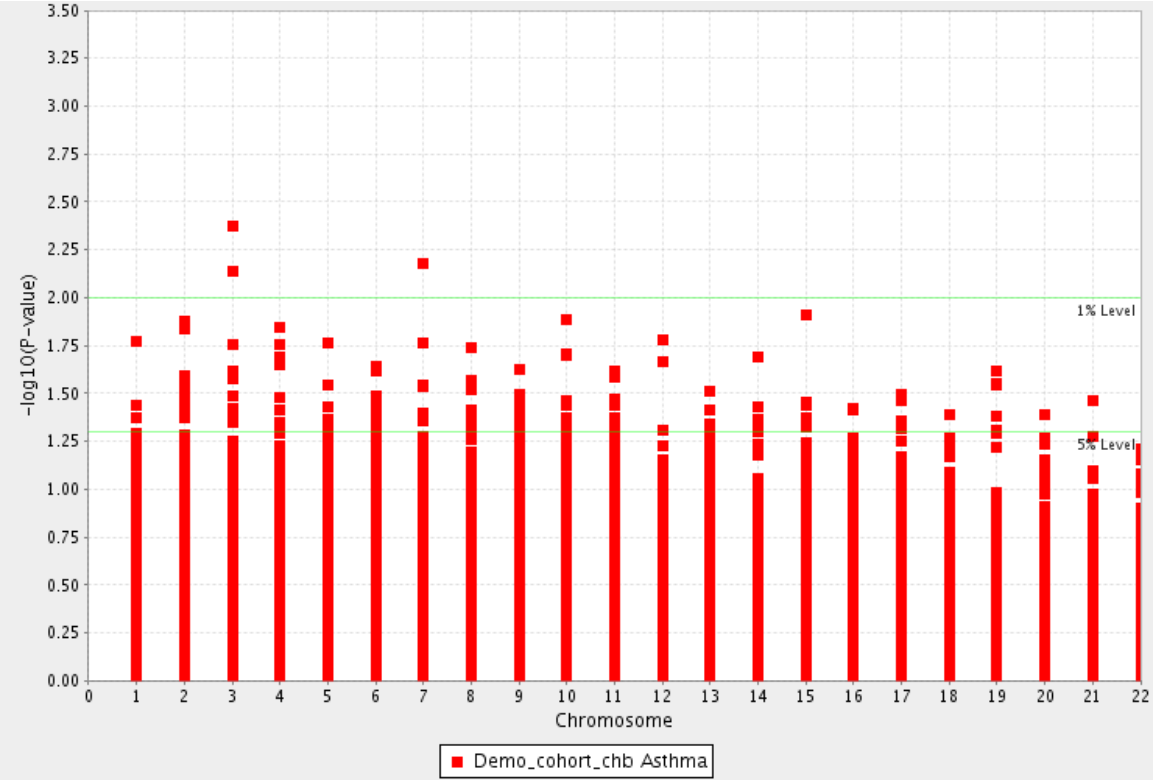


To help you determine associations, you are have to choose cohorts to do your association with. If you want to select multiple cohorts, you can hold down the CTRL key and then click on each cohort of interest.

Then, you will need to choose which phenotypes you are most interested in looking at. For the practice dataset there is only a dummy "asthma" phenotype. If you want to select multiple cohorts, you can hold down the CTRL key and then click on phenotype of interest.



Next, are some options that will help you narrow down your associations. You can specify p-value cutoff, SNPs, genes, and pathways.



Results

The association results page displays the single SNP association results in a Manhattan style plot. Each point on the plot has a mouse over with additional information. If a point is clicked on it will take the user to the SNP Summary page for the SNP that generated the association result. The number of associations that are returned is limited so that the page rendering is faster. If you feel like you would like to see more results, please contact your administrator.

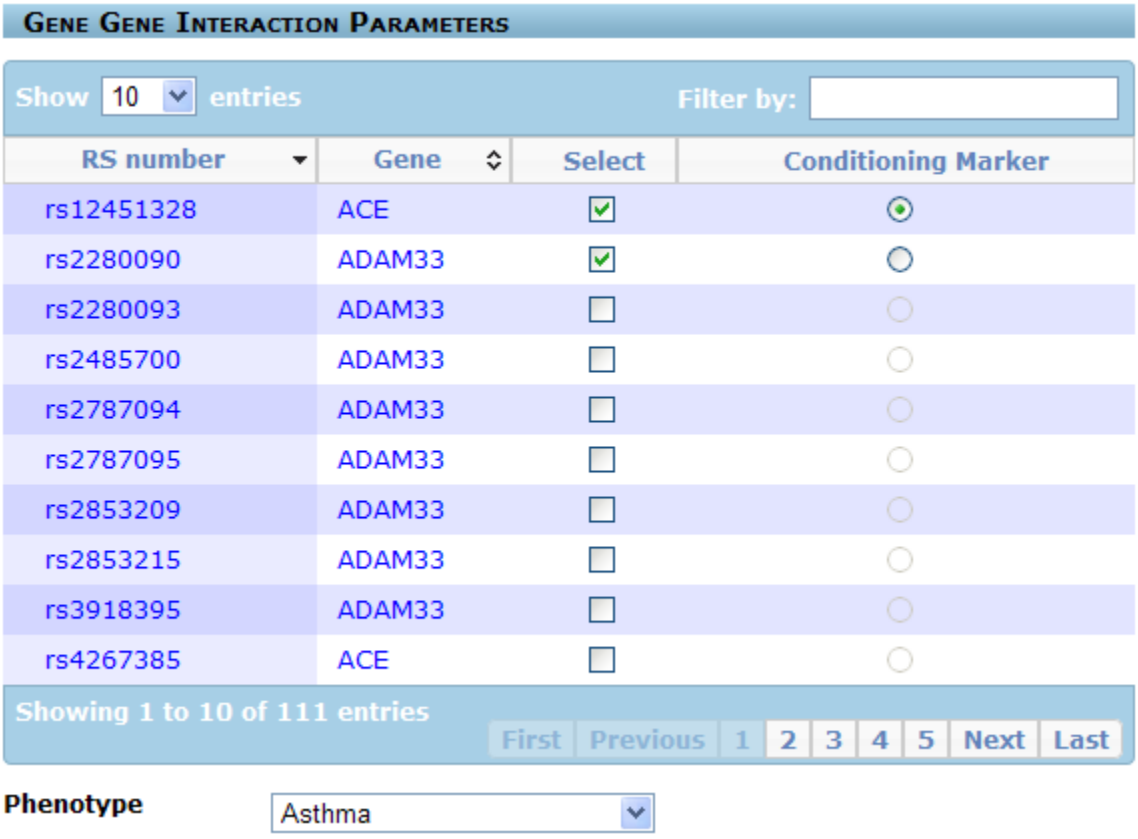
MAKE PATHWAY

The make pathway page allows you to create a custom pathway containing just the genes that you want to include. This will show how associated the pathway is to the selected phenotypes; the association is represented by Nyholt and Sydac corrected p-values. The custom pathway allows you to run a Gene-Gene Interaction test between two genes in their pathway.

GENE-GENE INTERACTION

The Gene-Gene Interaction uses Unphased (http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/) to test for interaction between SNPs in two

genes. From either the Custom Pathway or KEGG Pathway pages you have to select two genes from the genes list and then select “Gene-Gene Interaction”.



To use the Gene-Gene interaction, you will need to choose two SNPs in total, one from each of the 2 two genes. One of these SNPs has to be the conditioning marker. You will also need to specify which phenotype you want to test for.

MY RESULTS

This page shows the progress of requests that were made to the server, such as generating an analysis file or exporting raw genotypes.

SITE MAP

This page shows an overview of all the pages in Investigator Services. It allows you to quickly navigate to different parts of the application.

ADMIN SETTINGS

These options were implemented with the goal to make maintaining the database simpler. This page is only available to users with full permissions to use Investigator Services.

SHOW/HIDE SNPS/GENES

This option allows you specify SNPs and genes in the "Private" set (default) to another set in the database. Genes and SNPs that need to be copied can be inputted through a multiline textbox. To enter more than one gene/SNP, type one per line. Genes and SNPs in the Public set are available for viewing by the public.

UPDATE GENE LINKS

This feature makes sure that links are working and directing the users to the correct places. Gene links used on the gene summary page and redirect users to external resources for the gene, for example, NCBI.

UPDATE GENE ALIAS

A gene can be known by several different names. This feature updates the gene alias in your database to make searching for genes by their alias more accurate.

UPDATE KEGG PATHWAYS

The database stores a local copy of the KEGG pathways database. By running this option, the local copy is updated to reflect any changes that were made since the last update.